信息论笔记

宋佳欢

2019 年 9 月 9 日

目录

1 信息的度量

信息量 $I(x)=f\big(p(x)\big)$，函数 f 需满足下列四个条件:

1.f 单调递减，事件发生的概率越小，获得的信息量越大。

2. 当 $p(x)=1$，$f\big(p(x)\big)=0$

3. 当 $p(x)=0$，$f\big(p(x)\big)=\infty$

4. 两件独立事件同时发生的获取的信息之和为 $I(x,y)=I(x)+I(y)=f\big(p(x)\big)+f\big(p(y)\big)=f\big(p(x,y)\big)$

因此，$p(x,y)=p(x)p(y)$。根据这个关系，$I(x)$ 与 $p(x)$ 一定为对数关系。

根据上述四个条件可得:

$$I(x)=-logp(x)$$

其中负号是用来保证信息量是正数或者零。而 log 函数基的选择是任意的（信息论中基常常选择为 2，因此信息的单位为比特 bits，即信息需要的编码长度；而机器学习中基常常选择为自然常数，因此单位常常被称为奈特 nats；底数为 10，单位则为 Hart）。

$I(x)$ 也被称为随机变量 x 的自信息 (self-information)，<u>描述的是随机变量的某个事件发生所带来的信息量</u>。

现在假设一个发送者想传送一个随机变量的值给接收者。那么在这个过程中，他们传输的平均信息量可以通过求 $I(x)$ 关于概率分布 $p(x)$ 的期望求得，随机变量 X 的<u>信息熵</u>的定义:

$$H(X)=-\sum_{i=1}^{n}p(x_i)logp(x_i)$$

熵越大，随机变量的不确定性就越大。是对所有可能发生的事件产生的信息量的期望。

‘